RESPONSE TO RFI FOR A NATIONAL PRIVACY RESEARCH STRATEGY

INTERTRUST TECHNOLOGIES CORPORATION

Contact: David P. Maher, EVP and Chief Technology Officer | dpm@intertrust.com

TABLE OF CONTENTS

RESPONSE TO RFI FOR A NATIONAL PRIVACY RESEARCH STRATEGY

INTERTRUST TECHNOLOGIES CORPORATION

This report outlines Intertrust Technologies Corporation's response to the questions posed in an RFI for a National Privacy Research Strategy

Brief Background on Intertrust

- Intertrust (www.intertrust.com) is a private corporation oriented toward research and development in trusted distributed computing. In recent years, we have moved most of our resources toward projects involving personal data protection and big data technologies.
- Relevant research projects focus on privacy-based information management systems for Human Genomics and Personalized Medicine, Personal Private Networks, and Trusted Intermediary Technology that allow the data mining necessary to provide goods and services, while increasing consumer control over the use of their personal data.
- We have extensive experience in establishing standards for content protection, and have contributed to and cofounded several International standards bodies.
- We also strategically invest in "big data" companies that provide data governance and trusted intermediary services.
- Our activities have stimulated reflection on new frontiers in the legal, social, and economic aspects of personal data, where we explore ways to define the behavior and obligations of a new concept of personal data fiduciary. We examine how to provide consumers with greater abilities to manage their personal data, to efficiently extract value from it, and to rebalance the power between consumers and commercial entities.

1) Privacy Objectives

We believe an overarching objective for privacy research is to develop practical means for protecting and governing personal data while at the same time ensuring the public and personal benefits of collecting and organizing that data.

We explain this objective in the context of three scenarios:

a) The use of personal data in the context of public health activities and personal health care, where information management architectures that use trusted intermediaries and data governance concepts can enable highly sensitive personal health and behavioral data to become increasingly useful without undue violations of privacy.

b) The use of personal data in a commercial context where personalized information services and content services with proxy players (providing consumers with free content) can thrive while obviating the need for consumers to be tracked or even to divulge information about themselves except to trusted data fiduciaries.

c) The establishment of personal networks of things that are Internet connected. As all kinds of objects on our person, in our homes, and in our vehicles become Internet connected, both the controls for those things and the data they collect from their many (and increasingly invasive) sensors need to be governed in practical and intuitive ways. Legacy secure networking technologies are far too complicated — even professional network administrators can hardly manage the various risks. New ideas for developing protective structures, governing access to shared resources, and delegating authority to others need to be developed and standardized. Otherwise, the complexity of managing these new and potentially useful Internet connected devices will become overwhelming.

Scenario 1: Public and Personal Health

Information about personal health is arguably the most privacy-sensitive information collected about individuals. In many other areas, consumers have demonstrated a willingness to reveal personal information for free services that provide value to them; this has generally not been the case with healthcare data. This is not to imply that healthcare consumers are privacy absolutists. In fact, studies have shown that patients understand and value information sharing when it is used to provide better healthcare [LEM]. The fundamental challenge, and the focus of our research in this area, is to understand how to make information available and useful to those who can use it in the service of better healthcare while respecting the privacy rights of individuals.

We believe that the dual goals of access and privacy are not mutually exclusive. The technology that we are developing as part of our research program allows maximum flexibility in access to sensitive information, consistent with security and privacy policies determined by the data stakeholders. We will explain the approach by describing three use cases: (a) the collection and use of genomic information, (b) medical use of self-reported or user-generated data, and (c) privacy-preserving access to information for public health and epidemiology.

Scenario 2: Personal Data in a Commercial Context

People like to get content and services for free, and recognize that proxy payers (usually advertisers) pay the way. With more content being delivered via Internet, and especially to mobile clients, targeting ads and personalizing recommendations can be efficient for both the consumer (because they are more relevant) and the advertiser or service provider (because they reach the right audience). The opportunity for this kind of targeting has spawned an industry for "trackers" who literally lurk online and derive profiles of people from their online behavior, primarily using various kinds of cookies. While some consumers may be vaguely aware of such tracking, it occurs beyond their practical control [WSJ], and while anti-tracking technology has been introduced, it is not proving effective. Tracking on mobile devices does not work in exactly the same way, but motivation to track consumers is even greater, as mobile devices include numerous sensors that can track location, online behavior, medical information, measurements, and physical activities. Tracking takes on a whole new meaning here, and the potential for massive, uncontrolled invasion of privacy is undeniable.

Can consumers be profiled for personalized ads and recommendations without proliferating the distribution of such personal information? Can consumers practically and conveniently control their profiles so that they cannot be used to discriminate against them? And most importantly, since the gross privacy violations are motivated by financial interests, can the commercial goal of efficiently reaching specific audiences be satisfied without the need for surreptitious or even partially coerced profiling? As a result of our research, we believe the answer to these questions is yes, and we will discuss some of the possibilities in the architecture discussions below.

Another aspect of profiling that is important to address in this scenario is scope. Profiling processes can be limited in scope and restricted to specific applications. Scoping restrictions allow individuals and institutions to enforce "responsible use", a concept highlighted in [BIG]. For example, profiling processes that are aimed at making content or product recommendations or personalized advertising can limit themselves to specific types of data and inferences from that data — principally inferences about demographics, interests, and intents, as well as some aggregated location data. Responsible use would demand discarding raw data. There are data architectures that can be employed that obviate the need for associating most (if not all) personally identifying information with the profiles, though that alone does not assure true anonymity. Data storage and processing approaches for this scenario should be able to address this issue. It would be useful to develop practical and efficient ways for individuals to curate their data and discard inferences that they do not care to be used, and to directly determine what services can use the remaining data.

The practice of having businesses collect and amalgamate data, and then to sell it and physically distribute it to third parties for undetermined uses encourages personal data trackers to collect as much data as possible about each person. Furthermore, it discourages the practice of discarding potentially irrelevant data, since it is typically not known *a priori* what data is going to be useful. As the amount of data that can be collected increases with the number of sensors deployed — and we are talking about trillions of sensors in the near future — uncontrolled data collection about individuals will be relentless. We advocate a model in which individuals own their data, have the ability to curate it, and can safely make it available in an informed exchange of value with entities who can monetize it without dispersing it. In the architecture section below we explain how these latter entities can apply that data to satisfy the interests of advertisers who want to reach specialized audiences. This would address the needs of the commercial scenarios associated with proxy payers, while preventing the dissemination of data into environments beyond the control of individuals.

Scenario 3: Private Personal Spaces and Ubiquitous Connectivity

Practically every physical object that a person might interact with can become Internet connected. Connections for devices and appliances in our home and vehicles provide means for both remotely controlling those devices and for collecting data from them.

We benefit from both the data and the ability to control devices remotely, as well as the ability to delegate to others the ability to collect info from our devices and to control our appliances. However, the data collected can be intensely personal. For example, an ordinary looking drinking glass [VES] can now monitor and report on what we drink, providing precise measurements on ingredients, distinguishing between brands of beverages, and measuring alcohol content, among other feats. Other Internet connected devices can monitor our digestion and metabolism, and still other devices can monitor our waste, all within the home. There are a number of benefits that can accrue, at least for certain people by using such devices. But this is just a tiny example. Privacy, Safety, and Security issues quickly arise from the ability to control and monitor windows, doors, lights, appliances, health monitors, etc. We believe that this scenario presents enormous challenges, and we are concerned that traditional methods at providing security for network-connected devices will be incorrectly deemed sufficient to address these challenges.

Traditional network security has already been proven to be tremendously difficult for even professionals to manage. Making changes to network connectivity and access controls can have consequences that professionals may be able to handle. We believe that the so-called Personal Internet of Things will be much more complex. People will need to delegate access to others in some natural way while still understanding the consequences of delegation and how it interacts with automation. Network security and access control has never been simple and intuitive, yet we'll need to find solutions for this scenario. Maintenance of privacy will be aligned with safety and security considerations. Taking into account the age and maturity of family members and friends make the risk scenarios even more complex.

In fact, it may be nearly impossible to make it easy and intuitive for typical consumers to administer access to and control of things that we normally view as mostly passive. Thus, we may need to provide configuration tools, perhaps with the help of computational capabilities and visualizations. Experimentation and research in user experience and user interfaces will be necessary to approach this area properly.

Finally, in this scenario we again recognize that there are benefits to society in general if we can find ways to use some of this personal data responsibly. The technology permitting absorption of the information into huge databases already exists, along with the ability to analyze it. This data holds the potential to yield insights into energy usage, building efficiency, environmental stresses, public health, etc. We see the overlap of concerns between this scenario and others where data in the aggregate is useful for the public, yet access to or leakage of the pieces is vexing. Much of our research is aimed at addressing these and similar problems.

2) Privacy Concepts

In the early days of the Internet people who “went online” felt a certain amount of exhilaration from the ability to communicate with anyone. The reality of identity theft and other online dangers had not yet reached the mainstream. People naively shared too much of their personal information and misinterpreted the malicious intent of bad actors towards their “small data”. Today the Internet is being used to automate a wide range of functions from the digitization of our personal and financial records to the collection and processing of information from trillions of sensors. In this new environment, different aspects of privacy need to be more closely examined. Instead of focusing solely on how much information we explicitly share with friends in our social networks, we must now shift our focus to our ability to control to whom and in what context information about us is collected and revealed by others, especially as many more automated capabilities are developed for collecting and disseminating our personal information from the hundreds of sensors we encounter every day.

The following list of concepts can help us to better understand the challenges to personal privacy, and ways in which we can meet those challenges:

Personal Data Value

One of the great realizations on the web is that personal data has value. Personal data can be used to measure and predict the chance that someone will spend money or act in a certain way. The web has created a new financial instrument, personal data, which has no regulations, oversight, or transparency in the market.

Personal Data Dispersion

How much and how far is our personal data dispersed when it leaves our direct control? When we share data with others or when it is collected from us without our immediate and direct knowledge, how widely is it dispersed? Could there be a method for tracking personal data as it travels across the web from entity to entity? The web has made it extremely easy to transfer and copy information. Digital pirating of content has disrupted the entertainment industry. The same principles apply to personal data, as it also has value, and can be easy pirated and sold to advertisers, or worse. On the one hand, easing of the flow of information greatly improves convenience and efficiency, while on the other it creates a higher risk to privacy. A requirement for a more privacy-conscious web in light of the three scenarios above might involve data tagging and tracking for personal data, time-controlled data sharing, and safe and secure digital personal lockboxes.

Data Ownership

Currently, most Internet companies thrive on a business model based on trading online services such as search or social networking in exchange for collecting personal data. Companies have crafted privacy agreements that give them ownership of any personal data collected on their website, including metadata, which they then store in perpetuity. There is no way to measure or calculate whether this is an equitable trade. These companies then sell their customer profiles to third parties for monetization in the form of further profiling or advertising. Consequently, an individual's data passes from the company s/he engaged with to other organizations without the individual's knowledge. This creates serious risks and vulnerabilities to an individual's right to data privacy.

Information leakage

In cases where we choose to govern our personal data yet continue to interact with services such as search and inquiry, how much data about us is leaked? This is an important concept in the context of anonymization. Due to current data ownership structures on the web, individuals have little control over or insight into information leakage.

Predictions and Insights Violating Privacy

With the near-ubiquitous sensors and tracking tools in our applications and devices, each action and activity creates more data. People might happily value a tool that tracks their activity. However, algorithms can track patterns across these small data sets to create metadata, predictions, and insights that are unexpected and violate privacy. For instance, the Target pregnancy prediction scandal [HIL], in which a young woman's purchasing records led the company to predict that she was pregnant and send her targeted offers that would appeal to a pregnant woman. While on the one hand, Target was more accurately marketing products that might be useful, the shopper had no knowledge and had not given the store consent to access her records, make predications about her personal life based on her shopping record, or send her targeted advertising based on the prediction.

Data concentration

While it might benefit us to keep all of our data in one place — a hard drive on our computer, or an old filing cabinet — this concentration of data poses huge risks to our security. In the physical world, we are able to ensure our privacy by locking our private documents in filing cabinets and closing doors and windows to safeguard against intruders. However, because digital data is spread across the web, little shards of private data are available across open and private networks, making it feasible for motivated parties to reassemble the fragments to uncover your full identity. While concentrating these accounts into one location, service, or tool might help to squash forgotten identity markers — that social network someone signed up for and forgot about; the credit card account they opened but never used —, this concentration also creates a more complete and valuable target for identity theft and larger privacy violations. While it may be embarrassing to have your social network account made public to more than just your friends or friends of friends, it can be devastating to have your bank accounts, medical records, and sensor data breached.

Control and Governance

Control and governance involve the mechanisms by which personal information is governed when it is made available to others, and by which individuals can control its use. At the moment, due to the lack of data concentration on the web and an imbalanced ownership structure of data, individuals have very little control over their own data. They might have access to it, but they have little control. Even if you delete an account, it is common practice for the company providing the service to store that data in their servers indefinitely. When a citizen engages in a service online, that citizen most often is relinquishing their control over their personal information in exchange for that service.

Trusted Intermediary

The Trusted Intermediary concept can be very effective when built into personal data services. Trusted intermediaries share common technologies with Digital Rights Management systems that have been used in the entertainment industry to maintain ownership and control over copyrighted content. A Trusted Intermediary can be used to make data useful and transferable without handing over complete ownership and control. Trusted Intermediaries are typically automated agents that can match interests between entities interested in finding an audience with specific attributes, and individuals that have those attributes. A Trusted Intermediary's role is to be a reliable, trustworthy agent for personal data. In many ways, Trusted Intermediaries are similar to financial brokers (such as eTrade or Scottrade) that individuals trust to handle their money and purchases. In this case, brokers never directly own the stock or the individual's funds; rather, they act as mutually trusted agents that connect buyers and sellers.

Personal Data Fiduciary

This is a new concept that is meant to provoke discussion on the obligations and liabilities of entities that help individuals govern and benefit from their data. These concepts are useful in discussing the delegation of authority and responsibility for managing data and providing useful and reliable services. This might be similar to regulations such as the Gramm-Leach-Bliley Act that regulates the financial industry, especially for fund managers and stockbrokers who manage an individual's wealth. If we recognize that, like our equities or financial products, personal data and privacy have value, then we need to regulate data brokers and traders and design structures to manage the industry that governs this wealth. We could look to the financial industry, which has philosophical similarities in that it deals with an amorphous, agreed upon value and trust put into a currency. The numbers in our bank accounts are just that, numbers, which are managed and in digital vaults that we trust to be safe and secure due to the protections and controls put in place to govern the industry. A similar structure needs to be put in place to govern privacy.

Transparency

This concept is important in the context of services that help people manage and govern their data. Transparency is high when it is straightforward to verify the safety and effectiveness of the means for protecting and using private data. A transparent system will be auditable for compliance with both statute and best practices, as well as with commercial claims for quality of risk abatement and policy management. A transparent system is one that can be tracked and evaluated, and in which an individual knows where his/her personal data has gone, who has it, what they have done with it, and what value has been created from it.

Balance of power

This concept is meant to illustrate the ways in which people are coerced to reveal information. Such coercion is often subtle. As we discuss such concepts of "opt-in" and "opt-out" and the possibilities for unfair discrimination, a notion of power-balance can be invoked when examining scenarios where data is demanded in return for services. At the moment, the balance of power on the web is highly skewed towards service providers. Individuals have little choice, other than to not use the web, which is not a fair or realistic option. Individuals should have the ability and the right to use the Internet without having to give up privacy.

3) Multi-disciplinary approaches:

Protecting data privacy straddles many disciplines, most notably the study of technology, law, and society.

We've discussed the massive automated collection, analysis, and dispersion of personal information. Addressing these privacy-violating practices will involve the automation of means for governing that data, including the provision of useful services that will help individuals cope with the complexity of management, and contending with the economic forces involved in seeking and using that data. In order to provide those services, and to make sure they are useful and effective, we will need to go far beyond purely technological solutions.

First, there are those things that border on the technological, including the user experience and user interfaces associated with data governance measures. We have seen a number of methods for data governance and policy negotiation fail due to inadequate consideration of the overall user experience. In short, it's a nuisance to do any extra work in order to be more secure and protect privacy. One of the reasons why people may seem little concerned about sharing their personal information on the web is that technology companies make it extremely easy to do so. To be able to counteract such privacy violations, the UI and the ease of use for improved privacy mechanisms will have to be automated, easy-to-use, and happen in the background without upsetting the immediate goals of the user.

The current healthcare system is full of inefficiencies and bad user interfaces. If we can make a more secure system that is also more useful and efficient, we can seize a great opportunity to move people towards more privacy-conscious systems.

Protecting privacy requires building regulatory frameworks for managing the brokering of personal data online. This regulatory structure does not exist at the moment, and as such, the market is extremely opaque. As discussed earlier, part of the solution is to have intermediaries provide services. Other parts of the solution can come from ensuring transparency in how that data is used. New approaches will take us to the frontiers of legal protection and remedies for both intentional violation of trust and for good faith behavior that, due to unanticipated events, can result in damages. We will need to define both the obligations and best practices of personal data fiduciaries, and make sure that the cost of risk management is aligned with the potential rewards for providing consumers with useful services.

Shining light onto this market by encouraging (or even mandating) Trusted Intermediaries, will greatly improve the ability to protect privacy online. The targeted advertising market is growing, and will continue to do so at record pace. Standard click-through rates on ads are roughly 1% at best. Google earned over USD$40b in revenue last year through ads that only work 1% of the time. This market has the potential to grow enormously, and yet there is almost no legal or regulatory structure for managing this trade of user's personal data for targeted advertising.

Healthcare is intrinsically multidisciplinary, involving not only patients and their medical caregivers, but also university researchers, public health officials, regulators, insurers, and a complex patchwork of laws that vary by state. Given the diversity of stakeholders with an interest in the data, the problem of simultaneously providing access to data and protecting the interests of all of the stakeholders becomes particularly challenging.

4) Architectures

The RFI requests descriptions of how privacy architectures will implement a "responsible use framework." The privacy architectures that we explore at Intertrust extend primarily from the concept of a Trusted Intermediary. With the TI approach, when personal data is collected, it always goes into a secure container. The data remains within the secure container while machine learning and data science and other analytic techniques are applied to make useful inferences about the data. The TI then provides results according to policies agreed to with the TI's stakeholders.

Big Data: Seizing Opportunities [BIG] describes many of the problems with attempts at anonymizing data. Specifically, there is overwhelming commercial pressure to re-identify and connect data fragments, overcoming those anonymization techniques, and they have been shown to be highly effective. By providing private, safe and effective means for commercial entities to find their audiences and engage with them, some of that commercial pressure is alleviated, but it will not disappear. Trusted Intermediary services can be designed to allow people to interact anonymously without having to expose any data, even so-called anonymized data.

The TI approaches has the advantage of simplifying high-level policies since the main idea is to keep data contained, and eliminate the need to directly reveal the data to anyone. The remaining focus is applied to Responsible Use concepts extended to what the trusted Intermediary does, and specifically it:

- Does what is implied with its agreements with its stakeholders
- Discards personal data that is not needed to carry out its mission and the aforementioned agreements
- Limits access and provides results to specifically identified parties per explicit policy
- Maintains the requisite security of the information containers
- Minimizes information leakage. This is distinguished from a break in security, as here we mean leaks that might result from disclosure of answers to queries that could include means for re-identifying people.
- Vets any externally provided programs that may operate on the data. This is especially critical in this context since one of the advantages of a TI approach is the ability to keep data isolated and immobile (so that it is not dispersed beyond the control of the TI and its stakeholders), and use constitutes the ability to bring computations to the data.
- Makes personal data curation and governance measures extremely easy and convenient for individuals to use through UX studies and superior UI design.

Below we discuss, for each of the three scenarios, how a TI approach works and how responsible use applies more specifically.

Scenario 1: Public and Personal Health

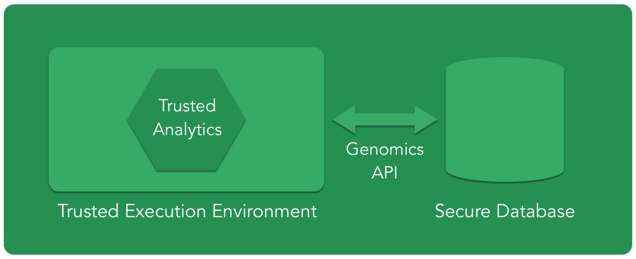
Using a Trusted Intermediary approach, we allow various actors to interrogate sensitive data via the intermediary of computer programs, rather than releasing the data directly to more- or less-trusted parties. These computer programs are sandboxed in a way that allows us to provide privacy assurances that would not be achievable with alternative architectures:

1. We can intercept calls between computer programs and the sensitive data set, allowing us to manage the interactions by:
    a. Authenticating the principal requesting that the program be run, the program itself, the programs authors and packagers, and others;
    b. Ensuring that the accesses performed by the programs happen in accordance with policies. The scope of these policies may be very broad, ranging from checking assertions about the program itself to requiring

that certain relationships exist between principals (e.g. that the accessing entity is a member of the care team for the patient that owns the data).

   c. Auditing all accesses for forensic purposes
   d. Potentially anonymizing data as it passes between a database and the program. For example, a cohort might be identified by an ephemeral ID rather than a list of individual IDs, thus obscuring the identities of patients in the cohort.
   e. Quantifying the information leaked by the interactions, possibly modulating access rights in response.

2. We can prevent the program from interacting with non-trusted services by controlling its network. For example, this approach limits the ability of programs to post sensitive data to untrusted third party sites.
3. Governance under this approach is responsive to the dynamic behavior of programs, not to assertions made by principals as to what programs will do, or how data will be used.



The notion of governing computations instead of attempting to govern human behavior has broad applicability in healthcare and in many other areas. The following examples illustrate how this approach might be used to empower various stakeholders in healthcare data.

Governing Genomic Data

According to recent estimates, 95% of all disease causing mutations are ‘rare’ — i.e., have an allele frequency of less than 0.5%. In other words, a cohort of 1000 would be expect to have fewer than one afflicted subject. Given the current economics of collecting whole genome sequences, few institutions have data sets large enough to provide any statistical power for studying rare diseases. Thus many institutions are working together through the Global Alliance for Genomics and Health to explore ways to foster data interoperability between institutions. Governance and privacy challenges arise when data sets are federated across policy boundaries. For example, some countries may not allow the export of raw genomic data, but may be comfortable with releasing statistics that cannot be used to identify research subjects.

Allowing trusted computations to traverse the policy boundaries — coupled with a secure execution infrastructure that guarantees that the computations are performed without interference — enables multiple stakeholders in vastly different policy environments to work together for research, clinical use, or public health.

Personal and Self-reported Data

Devices that collect healthcare information in a personal setting are become more and more prevalent. Despite their potentially lower accuracy, these devices have nearly continuous access to the patient, and thus may be clinically useful. Again, the question is how to get the relevant data into the hands of the people best qualified to interpret it, without overwhelming practitioners with data.

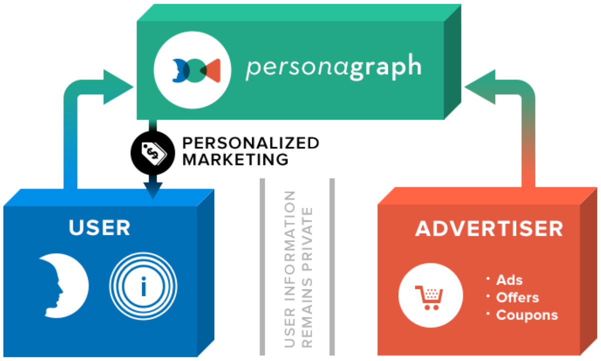
The computation-based approach is effective here as well. Stakeholders who have a specific interest in the data, such as deviation from the baseline a1c measurement for diabetics, can create computations that operate on the sensor data in a governed environment and finally act upon that data in a way that provides them with relevant, actionable, and relatively noise-free information. These computations can be performed according to polices specified by all of the various stakeholders in the data, including (and especially) the patient.

Data use for Epidemiology

The same techniques may be used by public health authorities for epidemiology and research. For example, suppose public health authorities wish to provide important information to individuals who either live in or who have traveled through certain regions. Software executing a governed environment might first identify those individuals (anonymously) by querying location data gathered by a personal agent and stored by a trusted intermediary. After identifying the individuals in question, the system can provide a notification specifying attributes such as urgency level and recommended actions. The individual can then decide which action, if any, to take. When fully scaled, such TI services can be a powerful means for both disseminating and gathering public health information. This approach can be extended for all manner of public health (and even private) research. For example, similar techniques can be used by researchers to discover candidates for medical research trials, to identify disease clusters, and the like.

Scenario 2: Personal Data in a Commercial Context

In this scenario, Personal Agents collect information about individuals and provide it to a Trusted Intermediary where associated trusted agents can either search for or listen to requests to match specific audience inquiries from public health, research, or commercial entities. Typically such inquiries include a template of scores on certain attributes. The template is matched against actual scores made from inferences of a given person's behavior. When a match is made, the agent can present a notification, advertisement, or recommendation to the individual that Personal Agent represents. This process is highly automated, and typically results in no leakage of individual data up to this point. When the individual responds to the notification, s/he can choose to either passively engage with the notifier where no information leakage need occur, or to engage with the notifier through secure channels, whereby the data policies of the notifier pertaining to any information leaked through the match will prevail.



When a Personal Agent searches for recommendations or appropriate advertisements or public health advisories, the agent can operate through an interface that combines search queries from millions of people emanating from a single IP address. This form of anonymization can be extremely effective, and can be done in a way that minimizes re-identification. The TI approach can make this more straightforward than similar techniques used in TOR [TOR] or Crowds [REI]. However, individuals must rely on the TI to protect the PIs operations, the data, and specifically the origin of requests.

Further information about this approach can be found in [PER].

Scenario 3: Private Personal Spaces and Ubiquitous Connectivity

The TI approach can be extremely powerful and there are methodologies for making it highly scalable and economical to operate. It is especially fruitful for scenarios a) and b) above, but it can also be used for scenario c) whereby a TI can help delegate access to a person's devices and sensor data, and it can employ analytics and graphical interfaces to help people understand the consequences of an access control decision.

In scenario 3) where we discuss the possibilities for personal private networks and ubiquitous connectivity, we mentioned the necessity of ensuring clarity of consequences when we delegate access to controls and data from our personal Internet connected devices. An intermediary service can perform the analysis of consequences and it can illustrate the consequences (or lack of them) in a convenient format. For example, if I make controls for a class of objects like windows, doors, and cabinets to a certain group of people, can those people inadvertently make those controls available to others? Are there safety consequences? If I allow a public interface to query my irrigation system status, what kind of data will be leaked, and to whom? While logically centralizing the data and remote control capabilities from all of your connected things has its risks, it can have significant advantages, as well, particularly with respect to aids in managing both data privacy and access privileges. A TI can provide individuals with state-of-the-art analytics, and it can provide convenient means for safely and securely granting access to precisely the right people among your family and friends.

The data and control interfaces for all of an individual's Internet connected things can form a logical personal data network, independent of physical networks. A data architecture that we call an Explicit Private Network (EPN) can focus on access controls and the governance of personal data associated with these things. In considering an architecture for this EPN, we use an approach similar to the TI approach, in which all data is sequestered and made unavailable and all access to controls are shut off except to a single individual. Then access is extended through highly explicit, intuitive, and convenient delegation protocols informed by analytics administered by intelligent agents in the cloud. That is, a cloud-based Personal Agent keeps track of all access control and delegation decisions, and aids in both analyzing and executing them. This approach, while it concentrates information in a cloud setting, can employ many classical data security, backup, delegation, and emergency exception measures. The data, including access control codes, can remain encrypted, yet crypto backup and escrow services [MAH] can be employed so that information is safely available when a person is incapacitated or dies. This is another role for a trusted data fiduciary.

The idea is to use Intelligent Personal Agents, operating in a secure environment, to provide data protection and governance services through a set of secure, rich, intuitive, convenient, and ubiquitous interfaces, providing each individual with performance aids that tame the complexity of governance and data protection. This can be done at the application layer without the necessity to interoperate with ineffective legacy network administrative controls and interfaces that have proven inadequate to the task. We can make use of open source and standards efforts to aid us to find superior methods and component architectures.

In the EPN concept, people (your family members) are nodes in one layer of the network, and your things are nodes in another layer, and external entities (for example public utilities) on another layer. Things can be grouped to form a compound objects with unified control and data access interfaces. When you acquire a new thing and it is connected to a physical network only one person has any access, but you can use a friendly Internet connected control panel to explicitly delegate various forms of access to other people. A TI provides performance aids for executing delegation instructions and making access conveniently available, and the consequences of such instructions are analyzed and explained. Information about people and things are stored in EPN databases to help in the analysis, so that people can be warned about the safety consequences of granting a child access to a dangerous device (either directly or indirectly through compound devices), and people can be clearly informed regarding possible information leaks affected by granting query capabilities to external entities.

References

[LEM] Lemke, A. A., Wolf, W. A., Hebert-Beirne, J., & Smith, M. E. (2010). *Public and biobank participant attitudes toward genetic research participation and data sharing. Public Health Genomics*, *13*(6), 368–377. doi:10.1159/000276767

[WSJ] The Wall Street Journal. *What They Know.* http://online.wsj.com/public/page/what-they-know-digital-privacy.html

[BIG] *Big Data: Seizing Opportunities, Preserving Values*, May 2014, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

[VES] Vessyl. https://www.myvessyl.com

[HIL] Kashmir Hill. *How Target Learned a Teen Girl was Pregnant before Her Father Did.* Forbes. 2/16/2012 http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

[TOR] The Onion Router. https://www.torproject.org

[REI] Michael K. Reiter and Aviel D. Rubin. 1998. *Crowds: anonymity for Web transactions. ACM Trans. Inf. Syst. Secur.* 1, 1 (November 1998), 66-92. DOI=10.1145/290163.290168 http://doi.acm.org/10.1145/290163.290168

[PER] Personagraph. *Privacy at a Crossroads.* http://www.personagraph.com/pg_privacy.pdf

[MAH] David Paul Maher. 1996. *Crypto backup and key escrow. Commun. ACM* 39, 3 (March 1996), 48-53. DOI=10.1145/227234.227241 http://doi.acm.org/10.1145/227234.227241